

FIGURES LEGEND

SUPPLEMENTARY MATERIAL

Figure S1. Evaluation of the stained cells quantification accuracy by the IS automated workflow vs. optical count (reference method). Bland–Altman plots of agreement between optical and automatic counts by the dedicated IS module implemented into the Developer XD software (Definiens, Germany) for 50 random tiles of colon cancer stained for CD3+ (left) and for CD8+ (right) are shown. Pearson correlation curves between counting methods are also presented.

Figure S2. Minimum surface (mm²) required for each marker (CD3+ and CD8+). Boxplots showing minimum area (mm²) required for each marker (CD3+ and CD8+), in each region (CT and IM) to estimate immune cells density equal to the whole region (+/- 10%) for each patient (n=538).

Figure S3. Inter-laboratory validation of the Immunoscore. Pearson correlation with *r* for mean percentiles (CD3+/CD8+) between two centers with the same number of cases (n=100) analyzed is shown. Inlay: contingency table showing the concordance of IS categories obtained for each case between two centers.

Figure S4. Evaluation of the Immunoscore as a prognostic biomarker. Kaplan Meier curves for time to recurrence (TTR) according to IS (IS 0-1, IS 2, and IS 3-4) in 448 patients with stage II-III colon cancer (**A**) and in 292 patients with stage II colon cancer (**B**). P-value was assessed by the log-rank test for trend. Hazard ratio forest plots for TTR, disease-free survival (DFS), and overall survival (OS) according to the

IS in stage II-III and stage II colon cancer patients (**C**). The P-value was assessed with the t-test log-rank (P) or the log-rank test for trend (P*).

Table S1. Evaluation of the Immunoscore inter-assay repeatability. Adjacent sections from three samples (S1, S2, S3) were cut to assess the densities of CD3+ and CD8+ T-cells (N1=44; N2=48; N3=36, respectively) and mean percentile of the IS. Quantifications obtained on adjacent slides for each staining were considered as duplicates. The contribution of one component (antibodies lots, instrument, revelation DAB kit lots, runs, or operators) was assessed between each duplicate for each sample. Differences of T-cells density (cells/mm²) or mean percentile (%) between two adjacent slides (duplicate) were calculated. Mean, standard deviation (SD), relative standard deviation (RSD), median, minimum, maximum, and the repeatability standard deviation (Sr) were assessed.

Table S2. Evaluation of the Immunoscore reproducibility. Adjacent sections from 3 samples (S1, S2, S3) were cut to assess mean percentile of the IS (N1=20; N2=24; N3=18, respectively). Contributions of variability of five components were performed with the ANOVA-Variance Component Analysis model by estimating variations between CD3+ and CD8+ antibodies lots (AB lots), instruments (Instr.), revelation DAB kit lots (Rev.), runs, and operators (OP.). Standard deviation (SD) and relative standard deviation (RSD) of mean percentile were calculated.